Integration of Apache Mesos over GPUs resources in the DEEP Hybrid DataCloud project

Aida Palacio Hoz[1], Marica Antonacci[2], Álvaro López García[1]
[1]Instituto de Física de Cantabria (UC-CSIC)
[2]Istituto Nazionale di Fisica Nucleare (INFN), Bari
Main contact: aidaph@ifca.unican.es

Introduction

Deep-Hybrid-DataCloud (DEEP) project aims to cope with the necessity to support several intensive computing techniques on specialized hardware, like HPC, GPUs or low latency interconnects. The project focuses on the integration of this specialized and expensive hardware under a Cloud Platform as OpenStack that can be used on-demand by researchers of different areas.

DEEP Problem

The project provides:

- **DEEPaaS:** API allowing to easily deploy user Deep Learning and Machine Learning applications
- **GPU** Cloud resources

How could we run DEEPaaS using GPUs?

How is the SOLUTION deployed?

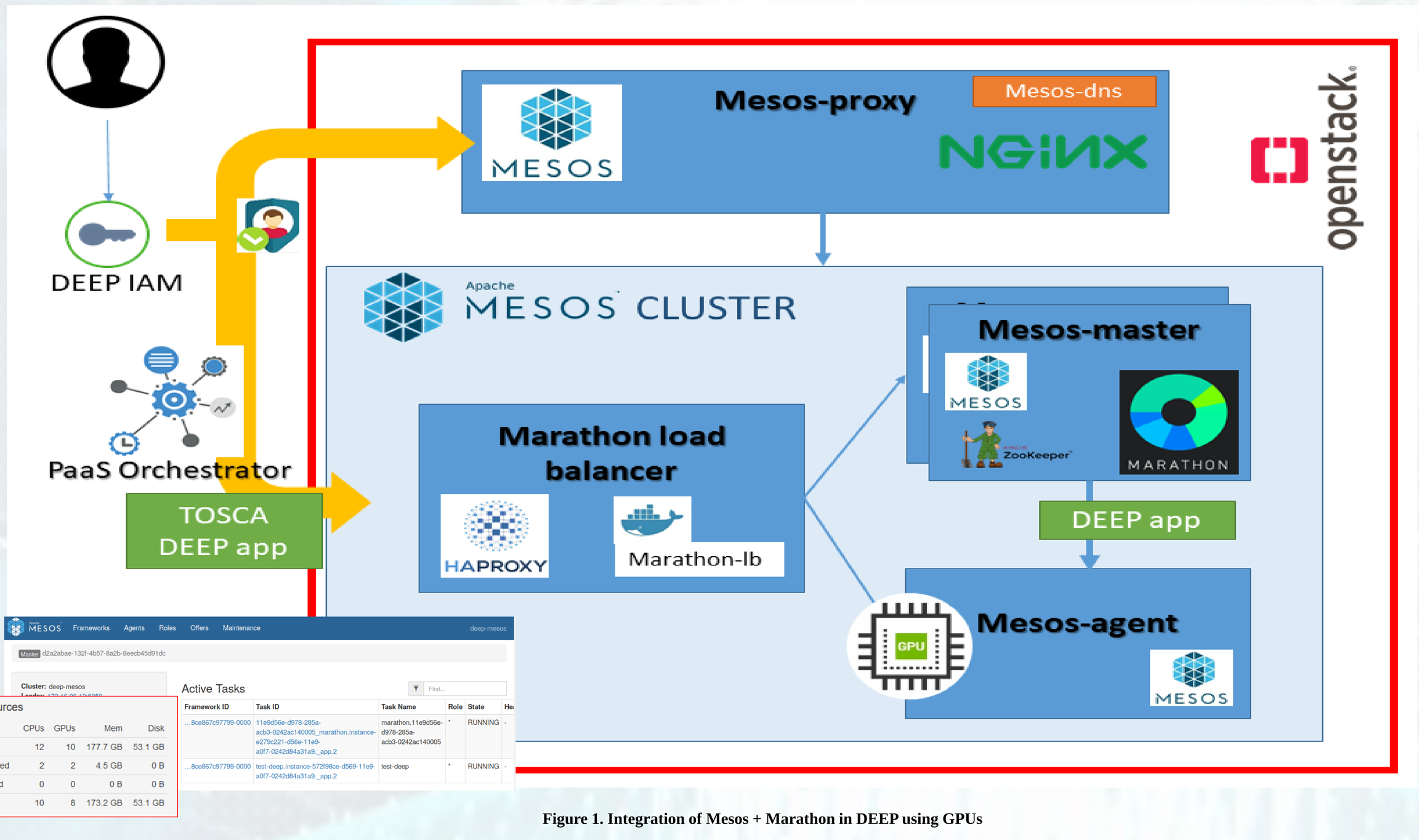


Figure 1. Integration of Mesos + Marathon in DEEP using GPUs

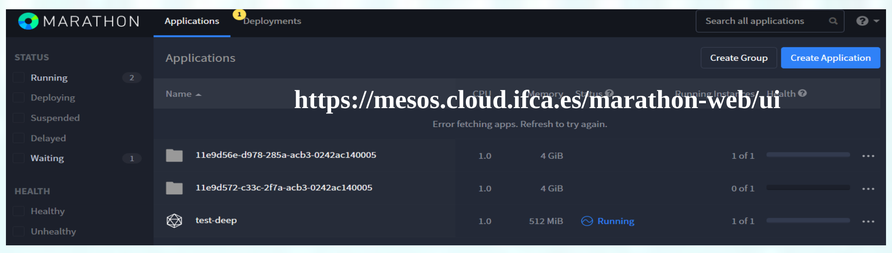
This solution consists in the integration of the following components within OpenStack:

- **Deep IAM**: Users authenticate before the processing
- **PaaS Orchestrator**: app deployment
- **Apache Mesos:** offers the resources for the processing
- **Marathon-lb:** Make the apps available to the outside

There are 2 ways for sending the application:

- Orchestrator
- Marathon web

Endpoints for orchestrator
https://mesos.cloud.ifca.es/marathon
https://mesos.cloud.ifca.es/mesos



Conclusions

By offering this solution to the community, DEEP provides a new way which users can deploy their applications using GPU resources. As well as this solution is properly integrated with the rest of the components of the project, the GPUs can be used on-demand in a well-managed way.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 777435.